DOCUMENT RESUME

ED 360 374 TM 020 278

AUTHOR Walsh, Lisa D.; Semb, George
TITLE Assessing Peer-Tutoring in the Classroom: A Comparison of Obtrusive and Unobtrusive Measures.
PUB DATE Apr 93
NOTE 73p.; Paper presented at the Annual Meeting of the American Educational Research Association (Atlanta, GA, April 12-16, 1993).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150) -- Tests/Evaluation Instruments (160)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *Classroom Observation Techniques; College Students; Comparative Testing; Feedback; Higher Education; Knowledge Level; *Measurement Techniques; *Observation; *Peer Teaching; Training; *Tutoring
IDENTIFIERS Obtrusive Measures; Self Report Measures; *Unobtrusive Measures

ABSTRACT

This study compared obtrusive observers and self-reports to a criterion, unobtrusive observers, to assess how accurately they measured tutor performance. Unobtrusive observers were used as a criterion for comparison because it was assumed that their measures were free of biases. Participants were 55 college-level peer tutors. Course sections were randomly assigned to one of four groups: (1) obtrusive observer and self-reports; (2) self-reports only; (3) obtrusive observer only; and (4) control (no obtrusive observer and no self-reports). All groups contained an unobtrusive observer. Tutors received training in tutoring skills and used those skills in a Personalized System of Instruction classroom to provide feedback to students. Tutors' skills were assessed before and after training. Obtrusive observers accurately assessed tutor performance, but self-reports were only accurate after training occurred. The self-report only group showed the least improvement on a test of written skills. No condition affected the amount of knowledge of course material. Six tables and six figures present study findings. Five appendixes contain the instruments used in the study. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED360374

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

✓ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

LISA D. WALSH

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ASSESSING PEER-TUTORING IN THE CLASSROOM: A COMPARISON OF OBTRUSIVE AND UNOBTRUSIVE MEASURES

Lisa D. Walsh

George Semb

University of Kansas

Paper presented at the national conference of the American Educational Research Association in Atlanta, April 12-16, 1993. Submitted to the Special Interest Group on Cooperative Learning: Theory, Research, and Practice, (Session #14.29). Prepared March 10, 1993.

BEST COPY AVAILABLE

Abstract

This study compared obtrusive observers and self-reports to a criterion, unobtrusive observers, to assess how accurately they measured tutor performance. Unobtrusive observers were used as a criterion for comparison because it was assumed that their measures are free of biases. Participants were 55 college-level peer tutors. Course sections were randomly assigned to one of four groups: obtrusive observer and self-reports; obtrusive observer only; self-reports only; and control (no obtrusive observer and no self-reports). All groups contained an unobtrusive observer. Tutors received training in tutoring skills and used those skills in a PSI classroom to provide feedback to students. Tutors' skills were assessed before and after training. Obtrusive observers accurately assessed tutor performance, but self-reports were only accurate after training occurred. The self-report only group showed the least improvement on a test of written skills. No condition affected the amount of knowledge of course material.

Table of Contents

Introduction 1
Method 9
Results 17
Discussion 36
References 44
Appendices 50
A. Unobtrusive Observer Evaluation 50
B. Effective Tutoring Skills Outline 51
C. Obtrusive Observer Evaluation 53
D. Tutor Self-Report 54
E. Written Skills Pre/Posttest 55
F. Coding Instructions for Written Skills Test 59

Assessing Peer-Tutoring in the Classroom: A Comparison of Obtrusive and Unobtrusive Measures

Businesses, group homes, hospitals, and schools regularly assess employee performance. To be useful, assessments must be valid and reflect performances that occur normally (in the absence of assessment). Three methods are frequently used to measure performance: obtrusive observers, unobtrusive observers, and self-reports (Kazdin, 1979).

Obtrusive observations

Participants typically know when obtrusive assessment occurs. A potential problem with this type of measure is reactivity (i.e. performance may change as a result of the observation) (Kazdin, 1979). Two studies demonstrated that reactivity can influence subjects' behavior. White (1977) found that the presence of an observer markedly reduced a family's "movement" in a lab that resembled a typical family living room. Observations from behind a one-way mirror revealed that the family covered more "distance" in the room than when no observer was present. In a study using unobtrusive observers to assess the effects of obtrusive observers, Brody, Stoneman, and Wheatrey (1984) also found that obtrusive observers affected play activity in preschool-aged and school-aged peer dyads (i.e. children engaged in more on-task play,

commanded or requested more behaviors from peers, and engaged in more social conversations, negative verbalizations, nonsense verbalizations, and task-related verbalizations).

Reactivity can result from variables other than the observation itself. For example, Ferris (1985) found that males performed better than females on electronic videogame tasks when an obtrusive observer was present. However, Berkey and Hoppe (1972) and Baron, Moore and Sanders (1978) found no gender differences in learning paired associations using obtrusive observers. Guerin (1986) also did not find gender differences in performance on a rotary-pursuit task when participants were observed obtrusively.

Observer movement also affects reactivity. Guerin and Innes (1982) suggest that reactivity occurs more when observers move quickly from one participant to the next. They also report that reactivity is greatest when the observer is in close proximity to the participant (e.g. when observers have to stand close to participants to assess their performance).

The way in which information is recorded may also produce reactivity (Laughlin & Wong-McCarthy, 1975). Laughlin & Wong-McCarthy (1975) found that an observer recording information impaired participants' performance on concept-attainment problems. However,

the observer's presence alone did not impair performance.

Jackson and Zedeck (1982), and Sanna and Shotland (1990) suggest that reactivity results from the individual's belief that observers possess some evaluative quality even if they are not recording information (e.g. an observer's evaluation may be used to determine whether or not a pay raise occurs, or to determine course grades in school settings). Jackson and Zedeck (1982) and Sanna and Shotland (1990) also suggest that the observer's influence depends on the anticipation of positive or negative evaluations. Sanna and Shotland (1990) found that participants performed memory recall tasks better with the expectation of a positive evaluation than when they expected negative evaluations.

Baron, et al. (1978) suggest that distraction may strengthen reactivity. They argue that the conflict between performing a task and the distraction of the observer may consume the individual's attention (i.e. the participant does not attend to the task because attention is divided between the task and the observer). Thus, performance should not be affected when the observer's presence is easily ignored.

In summary, while it is clear that obtrusive observers may be used to assess performance, results

from such studies should be interpreted with caution due to the potentially confounding effects of participant reactivity. Reactivity may be influenced by variables such as gender differences, observer proximity, whether the observer is performing an evaluative function, whether a positive or negative outcome is anticipated, and how distracting the observer is.

Self-reporting

Self-reporting, another method used to assess performance, is easy to administer (Howard & Daily, 1979; Mezoff, 1981). Bassett and Meyer (1969) also suggest that self-reporting may clarify procedures and responsibilities for the participant as well as provide feedback about their performance.

Reactivity may also occur with self-reporting because the participant is aware that assessment is occurring (i.e. observation of one's own behavior may influence performance) (Kazdin, 1979). In some situations, reactivity may be positive. For example, Lloyd, Bateman, Landrum, and Hallahan (1989) found that self-reporting produced improvements in arithmetic productivity and attention to task among special education children. Broden, Hall, and Mitts (1971) found that self-reporting increased the study time of one student and decreased another student's talking out

of turn.

Participants often rate themselves prior to and after training. However, results from pre/post measures may not accurately reflect participant performance. Howard and Daily (1979), Howard, Schmeck, and Bray (1979), Howard, (1980), and Mezoff (1981) argue that participants' conceptions about how they are to perform are inaccurate prior to training (e.g. participants may rate themselves high prior to training because they believe that they know all there is to know). However, after training, the individual's frame of reference may change. That is, they now have a new understanding of their performance and how to rate it. This change in frame of reference is referred to as a "response-shift." Evidence of response shifts in self-reporting are called "response-shift biases" (Howard & Daily, 1979; Howard, Schmeck, and Bray, 1979; Howard, 1980; and Mezoff, 1981). Because results derived from pre/post self-reporting may be confounded by response-shift biases, the effectiveness of training may be misstated.

In a literature review of 55 studies on self-evaluations, Mabe and West (1982) found a mean validity coefficient of 0.29. This low validity score indicates that either individuals tried to enhance their performances or lacked the ability to assess their own

performances objectively (i.e. individuals are too biased to produce reliable and valid recordings of their own behavior). McEnery and McEnery (1987) also found self-reporting to be more lenient than observer ratings. On the other hand, Heneman (1974); and Regan, Gosselink, Hubsch and Ulsch (1975) found that participants' self-reporting are more critical than independent observer ratings.

In summary, self-reports are easier to use and cost less than outside (obtrusive) observers. Self-reporting may also enhance performance as well as provide feedback. However, results from self-reporting studies may not be as accurate as other measures due to the possibility that individuals may not be able to accurately assess their own performances, particularly prior to training.

Unobtrusive observations

In the third method of assessment, unobtrusive observation, participants are not aware that their performances are being observed. An obvious advantage of the use of unobtrusive observers is that participant reactivity is not a factor. However, it is difficult to conduct unobtrusive observations without the participant's knowledge. There are also ethical issues to consider when using unobtrusive observers. One such issue is informed consent. When an experimenter asks

participants to give consent to participate in research, the experimenter is informing the participants that assessments will be taking place. Kazdin (1979) offers a solution to this problem by requesting participant's consent to several different types of assessment procedures. Therefore, the participants would not be aware of which assessment procedures were in effect. Another solution is to not inform participants of unobtrusive procedures until after they occur. Participants may then be debriefed and offered the option of whether the information obtained about their performances may be used for research purposes.

The present study was conducted in a tutoring environment that is similar to many training settings (e.g. companies train employees to serve customers, and hospitals train staff to interact with and care for patients). It is hypothesized (1) that observers will affect tutors' performance, (2) that tutors will not be able to rate their performances accurately prior to training, and (3) that they will be able to do so accurately after training.

Peer tutors were trained in tutoring skills and asked to use those skills in the classroom. The classroom consisted of 5 tutors, an instructor, and 40 college students. Students took quizzes and tutors

provided feedback on their performances. The three assessment methods described previously--obtrusive observers, self-reports, and unobtrusive observers--were used to assess interactions between the student and the tutor prior to and after training. The purpose of the study was to compare the obtrusive observers and the self-reports to the criterion, unobtrusive observers, to assess how accurately they measured tutor performance, and to determine if either of the obtrusive procedures altered tutor performance. Unobtrusive observers' measures were used as the criterion for comparison because it was assumed that these measures were free of biases.

Method

Participants

Participants were 55 undergraduate students, 14 males and 41 females, enrolled at the University of Kansas who had been chosen to be peer tutors for an introductory child development class. Each participant had completed the same course during a previous semester with a grade of an "A" or "B," and each voluntarily agreed to participate in the study by signing a university-approved research consent form.

Classroom procedures

The introductory child development course was taught using Keller's (1968) Personalized System of Instruction (PSI). Students took tests over 11 units at their own pace. After taking a test, tutors graded the test and gave feedback over incorrect items. In addition to grading tests, tutors maintained records of their students' performances. Each tutor was responsible for eight students.

Twelve sections of the introductory child development course were used in this study. Sections were randomly assigned to one of four groups: (1) obtrusive observer and self-reports, (2) obtrusive

observer only, (3) self-reports only, and (4) control (no obtrusive observer and no self-reports) (Table 1). All groups contained an unobtrusive observer. No group comprised more than 3 course sections. Four or five tutors were assigned to each section.

Obtrusive and unobtrusive observers

Four undergraduate students were recruited to serve as "unobtrusive" classroom observers. Throughout the study, their identities remained anonymous to the tutors. The "unobtrusive" observers behaved as students enrolled in the course. In addition to taking tests, they assessed tutors' performance across eight items using a 7-point Likert-scale data sheet (Appendix A). Items assessed effective tutoring skills such as approachability, listening to the student, diagnosing what the student knew, explaining and clarifying concepts, helping students to arrive at their own conclusions over missed questions, and overall feedback (Appendix B).

Two graduate students were recruited to serve as obtrusive observers. Their identities were made known to the tutors. Obtrusive observers also completed evaluations of tutors' performance using a similar 7-

Table 1

Experimental Groups and Measures

Groups	n	Measures
1. Obtrusive observer/ Self-report	15	Unobtrusive Obtrusive Self-reports
2. Obtrusive observer	13	Unobtrusive Obtrusive
3. Self-report	13	Unobtrusive Self-reports
4. Control (no obtrusive observer/no self-report)	14	Unobtrusive

point Likert-scale (Appendix C).

All observers were trained by the researcher to assess tutors on the following dimensions: knowledge of the material, listening, diagnosing, prompting, reinforcing, and approachability (Fitch, 1991). Training involved two sessions in which observers watched videotaped examples of student/tutor interactions that were previously recorded by expert student tutors. After each interaction, observers used evaluation forms that were identical to those used in the classroom to evaluate. Once the evaluation forms were completed, observers discussed differences in ratings and causes for disagreements. Observers also discussed ratings that were similar. Training continued until there was 100% agreement for each item.

Self-reports by tutors

Tutors completed a self-report of tutor skills during some tutor/student interactions. Self-reports consisted of eight items on a 7-point Likert-scale similar to the observer evaluation forms (Appendix D).

Tutor Training

All tutors were required to attend a peer tutoring seminar. In addition to tutoring in a classroom

section, the seminar included training in administrative procedures and tutoring skills, and reviewing course content. A 50 minute video-taped lecture on tutoring skills was shown on the first day of training. The video demonstrated desired behavior on: approachability; listening; diagnosing; clarifying and explaining concepts; prompting; and reinforcing (Appendix B). During the next two sessions, tutors watched previously recorded tutor/student interactions. Each video-taped session contained 4-6 interactions. On the final day of training, tutors role-played and practiced both student and tutor roles. When playing the role of a tutor, the tutor was asked to practice the skills described above. When serving as the student, the tutor was given a script of student responses. While observing video-taped and role-play interactions, the other tutors were asked to write down what the role-play tutor did well, and what they could have done to better approximate the models. After each interaction, tutors discussed and assessed the interaction with the instructor and each other.

Measures

At the beginning and at the end of the semester

tutors took an exam to measure knowledge of course content. The exam was similar to the final they had taken as students in the course the previous semester. It consisted of 50 multiple-choice and 10 fill-in-the-blank questions. There were four forms of the exam. Tutors received a different form than they had received as a student and also they received a different form at the beginning and end of the semester.

At the beginning of the semester and also following training, tutors were given a test requiring written responses to assess the acquisition of tutoring skills (Fitch, 1992 and Kuti, Hinton, Fitch, and Semb, 1992). The test consisted of two content questions and a tutor/student scenario (Appendix E). For the first two content questions, tutors were asked to describe what they would do if their students missed those questions. In the tutor/student scenario, the tutor was asked to describe what was done well and what they could have done differently in the scenario.

Experimental Procedures

The numbers of participants in the four groups are shown in Table 1.

Obtrusive and unobtrusive assessments and self-

reports were collected twice during the semester. The first set of assessments was collected during the first two weeks of the semester, prior to the time tutors received formal training on tutoring skills. Unobtrusive observers took quizzes and went to a different tutor each time to have their quizzes graded. After each interaction, unobtrusive observers went back to their desks and filled out evaluations of the tutor's performance. Each tutor was observed on at least two separate occasions by an unobtrusive observer.

Observers attended sections and stood behind seated students who were having their quizzes graded. This was done to insure they would be facing the tutors and listening to what the tutors said. During the interactions, observers would take notes to further enhance the obtrusiveness of the observation. At the end of the interaction, observers completed the evaluations based on what they had observed. Observers observed at least five different tutor/student interactions on each day data were collected. Prior to the observers' arrival in the classroom, tutors were told that someone would observe their performance, and

BEST COPY AVAILABLE

that the information collected would not be used to grade them, but rather to improve training.

During self-reporting, tutors filled out at least five self-reports on each day data were collected. Prior to the distribution of self-reports, tutors were informed that their responses would only be used to improve training.

The second evaluation period began immediately after tutor training was completed, approximately the fifth week of the semester. Data collection, again, consisted of four class days. Unobtrusive observers, obtrusive observers, and tutors followed the same procedures as described earlier.

Results

Reliability

An independent obtrusive observer assessed student/tutor interactions simultaneously with both the obtrusive and unobtrusive observers on 10% of all observation sessions. Ratings on each of the eight 7-point Likert-scale items for each such observation session were compared. An item's rating was considered an agreement when both the obtrusive or the unobtrusive observer and the independent observer provided the same point-value on the Likert-scale. The item's rating was a disagreement if the two numbers did not match. The number of agreements and disagreements on each item for all tutors was calculated. Interrater reliability was computed by dividing the number of agreements by the sum of agreements and disagreements. Interrater reliability on obtrusive observers' ratings was 80.0% during pretraining and 87.3% during posttraining. Interrater reliability on unobtrusive observers was 80.3% during pretraining and 90.0% during posttraining.

Two independent coders also rated units on the written skills test. A unit was the smallest segment that could be coded as a distinct behavior within a

category (Appendix F). Positive categories consisted of skills that tutors were trained to do (e.g. listening, diagnosing, prompting 1, prompting 2, reinforcing, approachability, and explaining), and negative categories consisted of those skills that tutors were trained not to do (not prompting, incorrect information, and not knowing information). Ratings of the two coders were compared. A rating was an agreement only if both coders rated it as belonging to the same category. The number of agreements and disagreements was calculated for all tutors. Interrater reliability was calculated by dividing the number of agreements by the sum of agreements and disagreements. Interrater reliability on test item responses was 99.7% for the pretest and 99.8% for the posttest.

Obtrusive and Unobtrusive Measures: Observer Group versus Control Group

Ratings on all Likert-scale items for each tutor were summed to obtain total scores for the obtrusive measures in the observer group and the unobtrusive measures in the control group during pre and posttraining conditions. These totals were summed

across tutors and divided by the number of tutors in each group to obtain mean scores. Means and standard deviations are reported in Table 2. Figure 1 presents means for the obtrusive observer group and the control group. A repeated measures analysis of variance by group (2) using pre- posttraining differences as a within-participants variable found no significant group effects, $F(1,25) = .02$, $p > .05$. The training differences were significant $F(1,25) = 35.82$, $p < .01$, but the interaction between groups and training was not significant, $F(1,25) = .93$, $p > .05$.

Self-Report Group versus Control Group: Unobtrusive Measures

Ratings on all Likert-scale items were summed to obtain total scores for the unobtrusive measures in the self-report group and the control group during pre and posttraining. Means and standard deviations are reported in Table 3. Figure 2 presents means for the self-report group and the control group. A repeated measures analysis of variance by group (2) using pre-posttraining differences as a within-participants variable found no significant differences between the two groups, $F(1,25) = .01$, $p > .05$. There was a

Table 2

Obtrusive versus Unobtrusive Measures

Measure	n		pre	post
Observer group				
	13	M	33.98	44.08
		SD	6.03	5.65
Control group				
	14	M	34.91	42.21
		SD	9.69	11.02

Figure 1 Pretest-posttest means: Obtrusive observers' mean score in the observer group versus unobtrusive observers' mean score in the control group

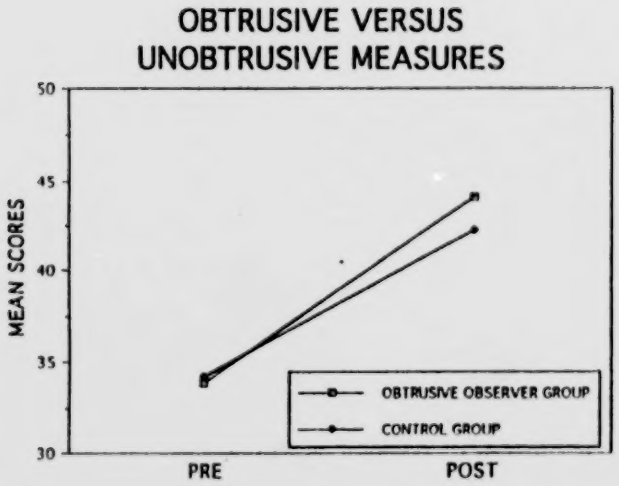
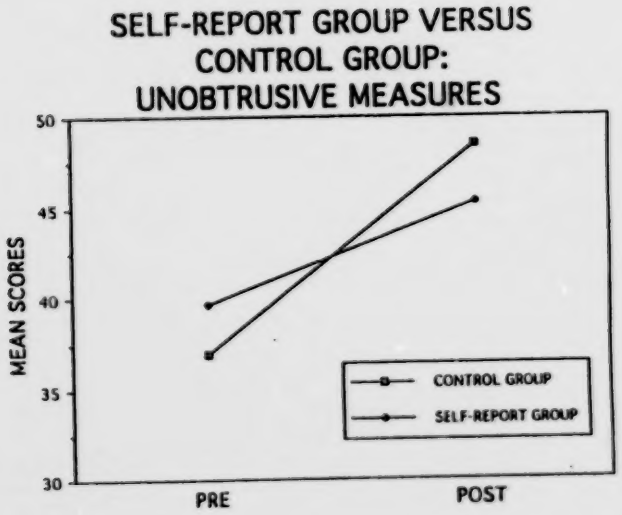


Table 3

Self-Report Group versus Control Group: Unobtrusive Measures

Measure	n		pre	post
Self-report group				
	13	M	37.00	48.50
		SD	9.85	4.63
Control group				
	14	M	39.67	45.37
		SD	6.76	6.82

Figure 2 Pretest-posttest means: Tutors' self-report mean score versus unobtrusive observers' mean score in the control group.



significant training effect, $F(1,25) = 30.11$, $p < .05$, but no significant interaction between groups and training, $F(1,25) = 3.42$, $p > .05$.

Observer/Self-Report Group versus Control Group: Unobtrusive Measures

Ratings on all items of the Likert-scale were summed to obtain total scores for the observer/self-report group and the control group. Means and standard deviations are reported in Table 4. Figure 3 presents the obtrusive/self-report group and the control group means. A repeated measures analysis of variance by group (2) using pre- posttraining differences as a within-participants variable found no significant group effects, $F(1,27) = .50$, $p > .05$. The training effect was significant, $F(1,27) = 36.05$, $p < .01$, but there was no significant interaction between groups and training, $F(1,27) = .19$, $p > .05$.

Tutor Self-Reports and Unobtrusive Measures: Self-Report Group

Ratings on all Likert-scale items were summed to obtain total scores for tutors in the self-report group and for unobtrusive measures collected for the same group. Means and standard deviations are reported in

Table 4

Obtrusive Observer/Self-Report Group versus Control Group: Unobtrusive Measures

Measure	n		pre	post
Obtrusive observer/ self-report group				
	15	M	37.40	44.67
		SD	7.19	6.36
Control group				
	14	M	37.00	48.46
		SD	9.85	4.68

BEST COPY AVAILABLE

Figure 3 Pretest-posttest means: Unobtrusive observers' mean score in the obtrusive observer/self-report group versus unobtrusive observers' mean score in the control group.

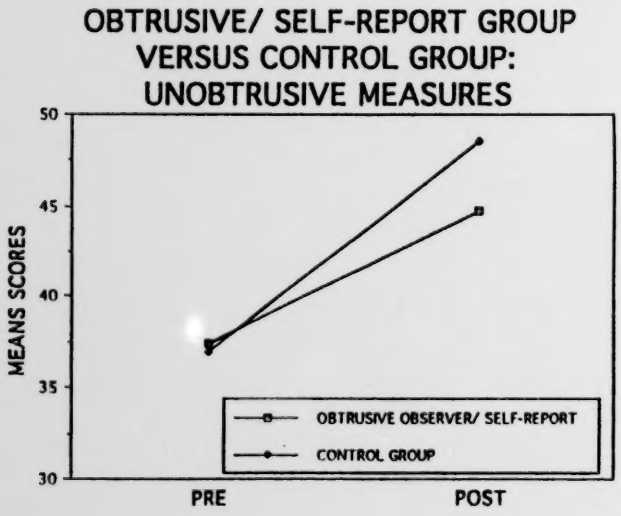


Table 5. Figure 4 presents tutor self-report means and unobtrusive observer means. A repeated measures analysis of variance by group (2) using pre-posttraining differences as a within-participants variable found significant main effects for group, $F(1,24) = 5.14$, $p < .05$, significant training differences, $F(1,24) = 20.00$, $p < .01$, and a significant interaction between groups and training, $F(1,24) = 7.77$, $p < .01$. A post-hoc Tukey was computed to identify which groups contributed to these differences. The analysis showed that the pretraining self-report mean was significantly higher than the pretraining unobtrusive mean ($p < .05$).

Knowledge of Course Content

Means were computed by summing tutors' raw scores on the course content exam for each group and dividing by the number of tutors in each group. Means and standard deviations are reported in Table 6. Figure 5 presents the groups means on the pre and posttest. A repeated measures analysis of variance by group (4) using pre and posttraining differences as a within-participants variable found no significant main group effect, $F(3,51) = 2.16$, $p > .05$, significant training

Table 5

Tutor Self-Reports versus Unobtrusive Reports

Measure	n		pre	post
Tutor self-reports				
	13	M	46.71	48.09
		SD	2.90	4.49
Unobtrusive reports				
	13	M	39.71	45.68
		SD	6.76	6.73

Figure 4 Pretest-posttest means: Tutors' self-report mean score versus unobtrusive observers' mean score in the self-report group.

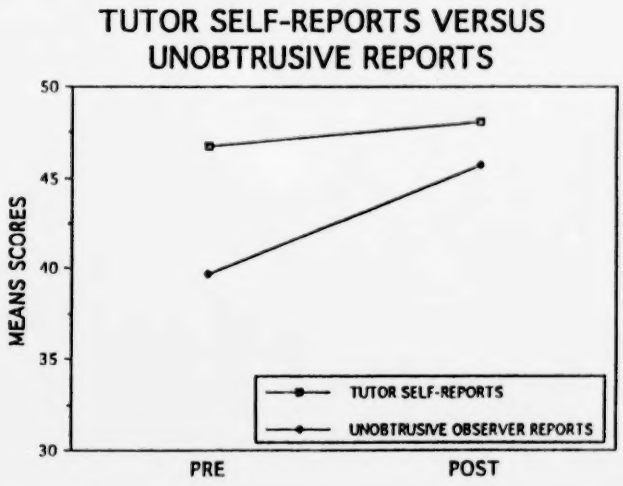
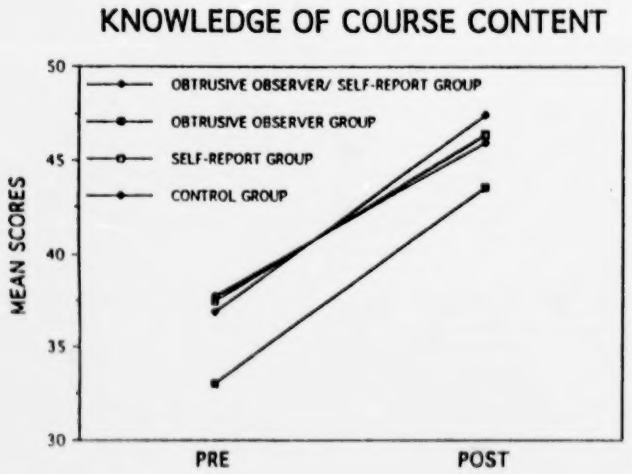


Table 6

Knowledge of Course Content

Group	n		Pre	Post
Self-report	13	M	37.43	46.36
		SD	6.48	4.58
Observer/self-report	15	M	36.85	47.46
		SD	4.12	4.14
Observer	13	M	33.00	43.53
		SD	5.32	8.05
Control	14	M	37.69	46.00
		SD	6.05	4.93

Figure 5 Pretest-posttest means: Knowledge of course content as a function of group membership.



differences $F(3,51) = 142.96$, $p < .01$, but no significant interaction between groups and training, $F(3,51) = .51$, $p > .05$.

Written Skills Test

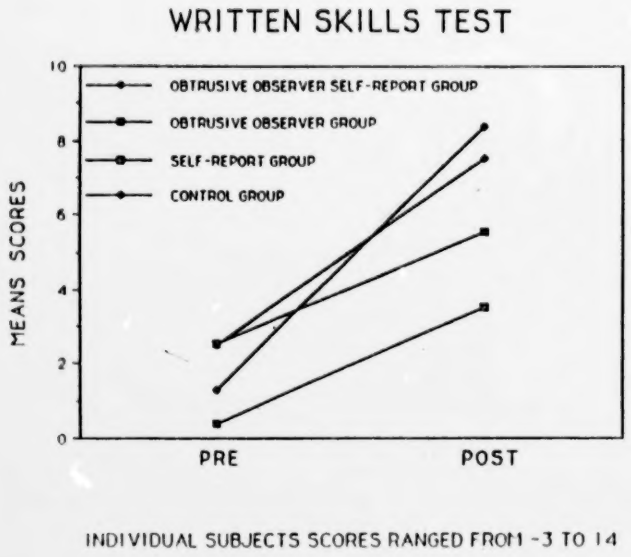
Scores for the written skills test were calculated by subtracting the sum of the positive categories (listening, diagnosing, prompting 1, prompting 2, reinforcing, approachability, and explaining) from the sum of the negative categories (not prompting, incorrect information, and not knowing information) to obtain a score for each participant for the pre and post written test. Scores ranged from -3 to 14. These scores were then added together for each group and divided by the number of tutors in each group to obtain a mean. Means and standard deviations are reported in Table 7. Figure 6 presents the groups means on the written skills tests. A repeated measures analysis of variance by group (4) with training differences as a within-participants variable found a significant main group effect, $F(3,51) = 5.88$, $p < .01$, significant differences from pre to posttraining, $F(3,51) = 63.98$, $p < .01$, and a significant interaction between groups and training, $F(3,51) = 2.78$, $p < .05$. A post-hoc

Table 7

Written Skills Test

Group	n		Pre	Post
Self-report	13	M	.38	3.54
		SD	2.43	4.48
Observer/self-report	15	M	2.48	7.50
		SD	2.03	3.11
Observer	13	M	2.54	5.54
		SD	3.35	3.21
Control	14	M	1.29	8.36
		SD	1.43	3.30

Figure 6 Pretest-posttest means: Written skills as a function of group membership.



INDIVIDUAL SUBJECTS SCORES RANGED FROM -3 TO 14

Tukey showed that posttraining self-reports were significantly lower than posttraining observer/self-report measures and posttraining control measures ($p < .05$).

Discussion

The results indicate that the presence of obtrusive observers did not affect tutors' performances in the classroom. Reactivity was not evident, suggesting that in this environment, obtrusive observers provide accurate assessments of tutor performance. The results also indicate that self-reports only accurately reflect performance when used after training. Prior to training, tutors rated themselves significantly higher than unobtrusive observers rated them. Results from comparisons of tutors' performances in the self-report group to tutors' performances in the control group indicate that performance did not improve when tutors used self-reports. Results from the pretraining written measure also indicate that there were no significant differences between groups. However, the self-report group's performance was significantly lower than the observer/self-report group and the control group on the posttraining written measure. Significant differences between groups were not found on material learned as measured by the posttest scores on the content exam.

In general, although self-reports may be easier to

administer and less costly, they appear to produce artificially high baselines. Thus, obtrusive observers may more accurately measure the effectiveness of training.

The results of this study suggest that the presence of obtrusive observers had a minimal effect on tutors' performance. This finding contradicts the hypothesis that predicted that observers would significantly affect tutors' performance. Differences may not have occurred prior to training, because tutors in all groups probably tried to tutor their best by "lecturing at" students (Fitch, 1992). However, observers were looking for skills that did not include lecturing at students. Differences were also not found after training. Several factors may have contributed to these results.

Baron, et al. (1978) suggested that reactivity results because observers distract the participant (i.e. participant has a conflict between performing the task and attending to the observer's presence). From this perspective, one might also consider the distracting stimuli present in the classroom. Not only do tutors attend to the students whom they are

tutoring, they also are aware of other students who need to take quizzes or study guides checked. Thus, the environment is "busy" with students entering and leaving the room throughout the class period as well as the instructor moving around the room to assist students and tutors. In this setting, the distracting qualities may be in the environment. Thus, the observer's presence would contribute very little to the distraction, and one would expect observer effects to be minimal.

From the perspective of Jackson and Zedeck (1982), and Sanna and Shotland (1990), reactivity results from the participant's expectations of the observer as an evaluator. However, in the present study, tutors may have learned to view other tutors and the classroom instructor as evaluators as well. During training, tutors evaluated each other and provided feedback on each other's performance. Furthermore, tutors sit next to each other in the classroom and have the opportunity to listen to each other while tutoring. In addition, the instructor often listens to tutor-student interactions. Thus, the presence of the observer in the classroom may not influence tutor performance.

Soskin and John (1963) as cited by Johnson and Bolstad (1975), and Purcell and Brady (1965) found that participants' reports of anxiety from observers disappeared after one or two days, suggesting that participants adapt to the observer's presence. This may also be true for the tutors in the present study. During role-play training, the trainer and other tutors evaluate each other. Therefore, tutors might have adapted to the presence of other tutors and observers as evaluators. Thus, reactivity to obtrusive observers in the classroom might be expected to be minimal.

Ceiling effects may also be a factor in explaining the lack of an observer effect. Tutors in the present study may have performed their best following training. Therefore, the presence of the observer could not increase performance. A more sensitive evaluation form may be needed.

The literature argues (Howard & Daily, 1979; Howard, et al., 1979; and Mezoff, 1981) that participants are too biased to estimate their own performances, particularly prior to training. This is based on the belief that individuals lack the appropriate frame of reference on which to assess performance. Furthermore,

this frame of reference changes after training resulting in a response-shift bias (Howard and Daily, 1979; Howard, et al., 1979; Howard, 1980; and Mezoff, 1981). Therefore, it was hypothesized that tutors' would be unable to accurately evaluate their own performance prior to training but would be able to do so after training. The present study's results confirm this hypothesis (i.e. a response-shift bias may have taken place in the tutors' understanding of skills).

Accurate self-reports may not have been only the result of response-shift biases. A number of characteristics are associated with accurate self-reports. These include ambition, sociability, approval, motivation, acceptance by others, higher intelligence, high achievement status, and internal locus of control (Mabe & West, 1982). Furthermore, individuals with higher intelligence and achievement status may better perceive their abilities, especially in terms of making more accurate judgements of their performances. In the present study, it should be noted that the individuals chosen to be tutors all earned a grade of an "A" or "B" in the same course taken the previous semester. Therefore, these individuals might

be considered to be of a higher achievement status compared with other students.

Results from the posttraining written measure for the self-report group were significantly lower than the control and observer/self-report group. These results suggest that self-reports may inhibit the understanding of what to do in the classroom. It may also be that the tutors in the self-report group believed that they knew everything they needed to know about tutoring. Therefore, they paid little attention to the skills taught during training, and were unable to describe what they should do specifically in a tutoring scenario.

Conclusion

The results of the present study indicate that obtrusive observers had no effect on tutors' performance. This may be the result of the setting in which the observations took place, the effects of the other tutors in the classroom, adaptation to observers during training, and/or ceiling effects.

The results also indicate that self-reports were not an effective method of assessment. Although they are easy to implement and less costly than other

assessment methods, self-reports may undermine the benefits because they are not valid measures of effectiveness of training as a result of response-shift biases.

Obtrusive observers also produce fewer problems than unobtrusive observers (e.g. ethical issues such as informed consent) because it is difficult to conduct unobtrusive observations without the participant's knowledge (Kazdin, 1979). Finding solutions for issues like informed consent creates further problems such as deciding how and when to tell participants that assessment procedures will be taking place without potentially influencing their performances.

Reliable and valid assessment procedures are crucial to organizations and individuals (Bunker, & Cohen, 1978), for predicting the success of training programs (Thorton & Byham, 1982), and for improving training effectiveness (Johnson & Sulzer-Azaroff, 1978). Based on the conclusions alone, obtrusive observers appear to be the most effective method to meet these needs when assessment procedures are to be conducted in settings similar to the one used in the present study. Research is needed to determine if the

results of this study can be applied to other settings (e.g. a cockpit) where participants perform with only one other individual present. Is it the environment that controls the tutors reactivity to observers or is it the training? More research is needed to answer that question.

References

Baron, R. S., Moore, D. L., & Sanders, G. S. (1978). Distraction as a source of drive in social facilitation research. Journal of Personality and Social Psychology, 36, 816-824.

Bassett, G. A., & Meyer, H. H. (1968). Performance appraisal based on self-review. Personnel Psychology, 21, 421-430.

Berkey, A. S., & Hoppe, R. A. (1972). The combined effect of audience and anxiety on paired-associate learning. Psychonomic Science, 29, 351-353.

Broden, M., Hall, R. V., and Mitts, B. (1971). The effect of self-recording on the classroom behavior of two eighth-grade students. Journal of Applied Behavior Analysis, 4, 191-199.

Brody, G. H., Stoneman, Z., & Wheaty, P. (1984). Peer interaction in the presence and absence of observers. Child Development, 55, 1425-1428.

Bunker, K. A., & Cohen, S. L. (1978, August). Evaluating organizational training efforts: Is ignorance really bliss? Training and Development Journal, pp. 4-11.

Ferris, G. R. (1985). Social influence and sex effects on task performance and task perceptions. Journal of Vocational Behavior, 26, 66-78.

Fitch, M. A. (1991). A coaching model of effective peer tutoring: Theory and research. Unpublished manuscript, University of Kansas, Lawrence, KS.

Fitch, M. A. (1992). Peer Tutor Training: A comparison of role playing and video evaluation for effects on student and tutor outcomes. Unpublished doctoral dissertation, University of Kansas, Lawrence, KS.

Guerin, B. (1986). The effects of mere presence on a motor task. The Journal of Social Psychology, 126(3), 399-401.

Guerin, B. J., & Innes, J. M. (1982). Social facilitation and social monitoring: A new look at Zajonc's mere presence hypothesis. British Journal of Social Psychology, 21, 7-18.

Heneman, H. G., III. (1974). Comparisons of self- and superior ratings of managerial performance. Journal of Applied Psychology, 59(5), 638-642.

Howard, G. (1980). Response-shift bias: A problem in evaluating interventions with pre/post self-reports. Evaluation Review, 4, 93-106.

BEST COPY

Howard, G. S., & Daily, P. R. (1979). Response-shift bias: A source of contamination of self-report measures. Journal of Applied Psychology, 64, 144-150.

Howard, G. S., Schmeck, R. R., & Bray, J. H. (1979). Internal invalidity in studies employing self-report instruments: A suggested remedy. Journal of Educational Measurement, 16(2), 129-135.

Jackson, S. E., & Zedeck, S. (1982). Explaining performance variability: Conditions of goal setting, task characteristics, and evaluation contexts. Journal of Applied Psychology, 67(6), 759-768.

Johnson, K. R., & Sulzer-Azaroff, B. (1978). An experimental analysis of proctor prompting behavior in a personalized instruction course. Journal of Personalized Instruction, 3, 122-130.

Johnson, S. M., & Bolstad, O. D. (1975). Reactivity to home observation: A comparison of audio recorded behavior with observers present or absent. Journal of Applied Behavior Analysis, 8, 181-185.

Kazdin, A. E. (1979). Unobtrusive measures in behavioral assessment. Journal of Applied Behavior Analysis, 12, 713-724.

Keller, F. S. (1968). "Goodbye, teacher ...". Journal of Behavioral Analysis, 1, 79-89.

Kuti, M. B., Hinton, J. A., Fitch, M. A., and Semb, G. B. (1992, April). Training peer tutors: The effects of video training and live role-playing. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Laughlin, P. R., & Wong-McCarthy, W. J. (1975). Social inhibition as a function of observation and recording of performance. Journal of Experimental Psychology, 11, 560-571.

Lloyd, J. W., Bateman, D. F., Landrum, T. J., and Hallahan, D. P. (1989). Self-recording of attention versus productivity. Journal of Applied Behavior Analysis, 22, 315-323.

Mabe, P. A., III., & West, S. G. (1982). Validity of self-evaluation of ability: A review and meta-analysis. Journal of Applied Psychology, 67(3), 280-296.

McEnery, J., & McEnery, J. M. (1987). Self-rating in management training needs assessment: A neglected opportunity? Journal of Occupational Psychology, 60, 49-60.

Mezoff, B. (1981, September). How to get accurate self-reports of training outcomes. Training and Development Journal, pp. 56-61.

Purcell, K., & Brady, K. (1965). Adaptation to the invasion of privacy: monitoring behavior with a miniature radio transmitter. Merrill-Palmer Quarterly, 12, 242-254.

Regan, J. W., Gosselink, H., Hubsch, J., & Ulsh, E. (1975). Do people have inflated views of their own ability? Journal of Personality and Social Psychology, 31(2), 295-301.

Sanna, L. J., & Shotland, R. L. (1990). Valence of anticipated evaluation and social facilitation. Journal of Experimental Social Psychology, 26, 82-92.

Soskin , W. F., & John, V. P. The study of spontaneous talk. In R. G. Barker (Ed.), The Stream of Behavior (pp. 228-281). New York: Appleton-Century-Crofts.

Thorton, G. C., III, & Byham, W. C. (1982). Assessment Centers and Managerial Performance. New York, NY: Academic Press.

White, G. D. (1977). The effects of observer presence on the activity level of families. Journal of Applied Behavior Analysis, 10, 734.

Appendix A

UNOBTRUSIVE OBSERVER EVALUATION OF TUTOR

Tutor:_____ Observer:_____
Date:_____ Class Day:_____ Time:_____ MWF TR

1. How **APPROACHABLE** was the tutor?

Distracted Grouchy Impatient	1	2	3	5	6	7	Eager to help Friendly Warm

2. How well did the tutor **LISTEN** to you?

Very poorly	1	2	3	5	6	7	Very well

3. How well did the tutor figure out what you knew and didn't know?

Very poorly	1	2	3	5	6	7	Very well

4. How well did the tutor let you explain things that you did not understand?

Not at all	1	2	3	5	6	7	Very well

5. When the tutor explained things, how accurate and relevant were the explanations?

Useless and Irrelevant	1	2	3	5	6	7	Useful and Relevant

6. What was the **QUALITY** of the **FEEDBACK** the tutor gave the you?

Poor	1	2	3	5	6	7	Excellent

7. How much did you learn from the interaction?

Not much	1	2	3	5	6	7	A great deal

8. Who did most of the talking?

You	1	2	3	5	6	7	Tutor

Appendix B

EFFECTIVE TUTORING SKILLS

What is a good tutor? As a tutor for HDFL 160, one of the skills you need is the ability to give effective feedback to your students. This requires listening, diagnosing students' problems with the material, asking questions, and reinforcing correct solutions, successful study habits, and good test performance. Essentially, a <u>good tutor "coaches" students how to learn the material for themselves</u>. Below is a list of effective tutoring skills, the behaviors involved in using the skills, and examples of these behaviors.

SKILLS	BEHAVIORS	EXAMPLES
1. <u>Listening</u>	eye contact; nodding paraphrasing what students say	"Uh-huh" "So you're confused about cohorts?"
2. <u>Diagnosing</u>	reviewing students folders; questions & statements about:	
	* students' progress	"May I see your study guide for Unit X?"
	* what they know	"Tell me everything you know about correlation and causation."
	* what they don't know	"What exactly don't you understand about this question?"
	* how they study	"Tell me how you study for this course. Do you read the entire chapter first?..."
	* explaining answers	"Why did you choose d?"

3. Prompting	questions and statements that:	
	• break material into components	"What happens during the first stage of child-birth?"
	• give examples	"Can you give me an example of an independent variable?"
	• ask students to explain or repeat in their own words	"Now, in your own words, tell me what I just said."
	• evaluate answers	"Which of these answers would you eliminate?"
	• suggest ways to think about the material	"How are you going to remember this concept?"
4. Clarifying	explaining, defining after diagnosing and prompting	"You're close - dependent variables are outcomes, independent variables are what differs between groups."
5. Reinforcing	smiling, encouragement, statements about correctness or quality of answer	"You're doing great!" "That's right." "Good example!"

1-20-92
M. A. Fitch

Appendix C

OBTRUSIVE OBSERVER EVALUATION OF TUTOR

Tutor:_____ Observer:_____
Date:_____ Class Day:_____ Time:_____ MWF TR

1. How **APPROACHABLE** was the tutor?

Distracted Grouchy Impatient	1	2	3	5	6	7	Eager to help Friendly Warm

2. How well did the tutor **LISTEN** to the student?

Very poorly	1	2	3	5	6	7	Very well

3. How well did the tutor figure out what the student knew and didn't know?

Very poorly	1	2	3	5	6	7	Very well

4. How well did the tutor let the student explain things that the student did not understand?

Not at all	1	2	3	5	6	7	Very well

5. When the tutor explained things, how accurate and relevant were the explanations?

Useless and Irrelevant	1	2	3	5	6	7	Useful and Relevant

6. What was the **QUALITY** of the **FEEDBACK** the tutor gave the student?

Poor	1	2	3	5	6	7	Excellent

7. How much did the student learn from the tutor/student interaction?

Not much	1	2	3	5	6	7	A great deal

8. Who did most of the talking?

Student	1	2	3	5	6	7	Tutor

BEST COPY AVAILABLE

Appendix D

TUTOR SELF-REPORT

Tutor:_____
Date:_____ Class Day:_____ Time:_____ MWF TR

1. How **APPROACHABLE** were you today?

Distracted Grouchy Impatient	1	2	3	5	6	7	Eager to help Friendly Warm

2. How well did you **LISTEN** to the student?

Very poorly	1	2	3	5	6	7	Very well

3. How well did you figure out what the student knew and didn't know?

Very poorly	1	2	3	5	6	7	Very well

4. How well did you let the student explain things that the student did not understand?

Not at all	1	2	3	5	6	7	Very well

5. When you explained things, how accurate and relevant were your explanations?

Useless and Irrelevant	1	2	3	5	6	7	Useful and Relevant

6. What was the **QUALITY** of the **FEEDBACK** that you gave the student?

Poor	1	2	3	5	6	7	Excellent

7. How much did the student learn from the tutor/student interaction with you?

Not much	1	2	3	5	6	7	A great deal

8. Who did most of the talking?

Student	1	2	3	5	6	7	Tutor (you)

Appendix E

WRITTEN SKILLS PRETEST

A. Read the following situations and write your response, as the student's tutor, in the space below.

1. Barry has missed the following question:

   T F Correlation is necessary for causation.

2. Michelle has failed the quiz for Unit 1 twice. She has missed questions about dependent and independent variables on both quizzes.

B. Evaluate the following tutor/student interactions. What does the tutor do well? What would you do differently if this were you student? Write your evaluation in the space provided. Use the back of the page, if necessary.

1. Tutor: (after grading quiz) You missed the question about correlation and causation again.

   Student: Really?! I thought I knew that!

   Tutor: Yeah, this is a tricky question. You just have to memorize the statement, "Correlation is necessary, but not sufficient for causation."

   Student: I did! And I can remember it as long as the quiz question asks it the same way. But when they change the words around, I get all confused.

   Tutor: To have causation, two things have to be correlated. But if two things are correlated, it doesn't necessarily mean that one causes the other. Does that make sense?

   Student: Yeah...I just hope that question isn't on the Review Exam.

   Tutor: Well, maybe you'll get it right next time.

WRITTEN SKILLS POSTTEST

A. Read the following situations and write your response, as the student's tutor, in the space below.

1. Connie has missed the following question:

   T F Reinforcement and punishment procedures are defined in terms of the effects they have on behavior and in terms of the perceived pleasantness or unpleasantness of the reinforcing or punishing stimuli.

2. Steve has failed the quiz for Unit 9 twice. He has missed questions about Piaget's concept of conservation on both quizzes.

B. Evaluate the following tutor/student interactions. What does the tutor do well? What would you do differently if this were you student? Write your evaluation in the space provided. Use the back of the page, if necessary.

1. Tutor: (after grading quiz) You missed the question about conservation.

   Student: Really?! What's the right answer?

   Tutor: You put "d"... the right answer is "c".

   Student: Why is "c" the right answer?

   Tutor: It's saying that changing the way something looks doesn't change how much of a substance is actually there. It has to do with the examples in the book about the balls of clay and the containers of water. Do you remember those?

   Student: Sort of...I just hope that question isn't on the Review Exam.

   Tutor: Well, maybe you'll get it right.

Appendix F

CODING FOR WRITTEN SKILLS MEASURES

Definition of a unit: The smallest segment that can be coded as a distinct behavior within a category. Adjacent phrases or sentences that are all related to the same topic or subject of a code are considered a single unit. **Separate units with a /.**

Category	Behaviors	Examples
L = Listening	* eye contact; nodding	"Uh-huh" "I see...OK"
	* paraphrasing what students say	"So you're confused about negative reinforcement?"
D = Diagnosing	questions or statements about:	
	* what they know	"Tell me what you know about correlation and causation."
		When going over the question, <u>before</u> eliminating any answers: "Do you know what the right answer might be?"
	* what they don't know	"What don't you understand about this question?"
		"Do you have any questions?"
	* how they study	"Tell me how you study for this course. Do you read the entire chapter first?"

Category	Behaviors	Examples
P1 = Prompting Level 1		
Tutor behaviors that do not involve the student, but may still promote learning. Primarily statements by the tutor that:		
	• give examples	"Prematurity is a good example to illustrate correlation."
	• suggest ways to students to think about the material	"Deductive logic goes from general to specific; inductive logic goes from specific to general. The letters "d" and "g" are closer together in the alphabet so general goes with deductive."
	• give study/test taking tips	"It's important to read the whole question."
		"You will need to know this for the review exam."
		"You should write the Study Guide in your own words."
	• use the text, SG, or TA to check themselves	"I'm going to look this up in the book so I don't confuse you."
	• <u>after diagnosing</u>, stress the importance of tutors using their own words to explain concepts	"If she still didn't understand, I would explain the concept in my own words."

BEST COPY AVAILABLE

Category	Behaviors	Examples
P2 = Prompting Level 2		
Tutor behaviors that involve the student in the learning, give the student an opportunity to respond, or require action on the part of the student. Primarily questions by the tutor that:		
	• break material into components	"What happens during the first stage of childbirth?"
	• ask for examples	"Can you give me an example of deductive reasoning?"
	• evaluate answers	"Which of these answers would you eliminate?"
		asking for the right answer <u>after</u> eliminating some options
	• ask students for ways they can think about the material	"How are you going to remember this concept?"
	• have the students use the SG or text	"Read the section about..."
		"Look at this chart in the text."
	• ask the student to read the question	"Why don't you read #4?"
R = Reinforcing		
	• smiling; encouragement after grading an exam	"You're doing great!"
	• statements about correctness or quality of answer	"That's right."
		"Good example!"

Category	Behaviors	Examples
A & P = Approachable and Personable		
	* using student's name	
	* encouragement before taking a test	"Good luck!" "Think positive!"
	* positive attitude or approach	"I'm sure you'll get it right on the next quiz."
	* offering help, generally	"Let me know if you have any questions, OK?"
E = Explaining		
	* clarifying, defining <u>after</u> diagnosing or giving the student a chance to respond	
NP = Not Prompting		
	Note: Code a statement/question as NP if the tutor did not use the opportunity to diagnose of ask the student questions, but simply began explaining. Code as E if the tutor is explaining in response to a student's question after determining what the student really doesn't know.	
	* giving answers rather than prompting	"The answer is b," in response to a student's request with no further elaboration.
		Launching into a lecture on a topic before finding out what the student already knows, or what the problem is.

Category	Behaviors	Examples
N = Negativity		
	* focusing on the student's failures too much	"Maybe you'll pass the quiz this time."
		"I can't believe you didn't know <u>that</u> one!"
	* blaming the tests, text, or SG for a student's poor performance	"Yeah, there are a lot of bad/trick questions on this quiz."

I = Incorrect information or explanation about course content or procedures, or **bad advice** (e.g. tells student to memorize conceptual material such as correlation or independent/dependent variable rather than focussing on understanding. If advice is about memorizing basic terms or stages, code as Prompting Level 1).

DK = Doesn't know information or is unprepared to help the student (e.g. tutors send students up to instructor or to another tutor for an explanation before attempting to answer questions themselves).

M. A. Fitch
5-23-90